Improving Efficiency and Coherence in Evolutionary Story Generation

Pablo Gervás
Facultad de Informática
Universidad Complutense de Madrid
Madrid, 28040, Spain

Abstract

A significant challenge for evolutionary approaches to story generation is to find a genetic representation for a story draft that allows mutation and crossover operations while also being able to capture the constraints of coherence between the different parts of the story. This may be achieved by defining a narrative draft in terms of combinations of knowledge structures that capture its structure. The present paper reviews a previously existing solution for the evolutionary generation of stories, both in terms of its representation, the evolutionary operators and the fitness function, and outlines an alternative solution that improves upon it. The two solutions are compared in terms of coverage of the search space, efficiency of the evolutionary search process, and quality of the resulting narratives.

Introduction

Evolutionary solutions have proven to be appropriate for implementing story generation systems based on models of desirable stories rather than on models of how humans build stories. This is because the essence of an evolutionary process lies in informed selection among a population of candidates, with the construction of the candidates being modelled on evolution via random mutation and crossover.

The challenge for applying this type of process to story generation arises from the choice of a genetic representation. A genetic representation for a story draft must allow mutation and crossover operations while also being able to capture the constraints of coherence between the different parts of the story. Solutions based on exclusively local representation of the different spans of the story will lead to outputs similar to those produced by the *exquisite corpse* technique of the Surrealists (Adamowicz 1998) or the *cut up* technique of the Dadaists (Cran 2013). Fragments cut out literally from different drafts will most often not make sense when put together in a new one. To avoid this problem, representations must be chosen that represent the structure of narrative in a way that captures its internal relations, but which, when some part of it is altered–as by evolutionary operators of mutation or cross over– it results in a different narrative that is also structurally coherent.

The present paper reviews a previously existing solution for the evolutionary generation of stories, both in terms of its representation, the evolutionary operators and the fitness function, and outlines an alternative solution that improves upon it. The two solutions are compared in terms of coverage of the search space, efficiency of the evolutionary search process, and quality of the resulting narratives.

Previous Work

The work presented in this paper requires understanding of three aspects of story generation: plot representation, prior evolutionary approaches and the existing approach used as starting point.

Basic Challenges of Plot Representation

Good stories have plot: the events in them are connected by a sense of causality (Forster 1927). Forster's famous argument states that "*The king died. The queen died.*" is a chronology of events, but "*The king died. The queen died of grief.*" is a plot. Knowledge-based procedures for story construction rely on capturing relations between events in some form in the representations they use for stories. Causal relations between events can be captured over complete *story schemas* (Booker 2004) or by defining smaller building blocks–such as *planning operators*–that define preconditions and postconditions with other elements in the story (Young et al. 2013). An intermediate approach relies on *axes of interest* or AoIs–small sequences of *plot atoms* representing events connected by plot-relevant causality and sharing characters in *roles* important to the plot (Gervás 2019). Table 1 shows an example of two AoIs combined into a simple plot.

Evolutionary Story Generation

Evolutionary algorithms have been applied to combine story fragments involving particular entities to the story, relying on a fitness function that combines coherence and interest of the story (McIntyre and Lapata 2010) or to generate small narrative fragments for text-based games using an evolutionary solution driven by novelty (Fredericks and DeVries 2021).

Other approaches have combined planning-based techniques to generate stories with evolutionary selection based on fitness functions. Aspects considered in the fitness functions are the believability of the story and the percentage of the user-defined goals the current story accomplishes (Kartal, Koenig, and Guy 2014) or degree of matching between

AoI	Plot Atom	Roles
ABDUCTION	Kidnapping	(abductor=x, abducted=y)
	Rescue	(abducted= , rescuer=z)
CALLTOACTION	Call	(called=hero,caller=sender)
	Reward	(rewarded=x)

(a) two axes of Interest (AoIs)

AB	Kidnapping(abductor=*villain*,abducted=*victim*)
CA	Call(called=***hero***,caller=*sender*)
AB	Rescue(abducted=*victim*,rescuer=***hero***)
CA	Reward(rewarded=***hero***)

(b) a combination of them into a simple plot (protagonist in Bold, rest of the characters in Italic).

Table 1: Plot representation in terms of AoIs.

AoIs	`Abduction`	(relation)	`CallToAction`
Shared roles	`hero`	=	`hero`
Sequencing	`Abduction`	<	`CallToAction`
	`Rescue`	>	`CallToAction`
	`Rescue`	<	`Reward`

Table 2: Example of constraint: the hero of both AoIs must be the same (line 2), the abduction must take place before the call to action (line 3), the rescue must take place after the call to action (line 4) and before the reward (line 5).

the tensions in the story and a target curve of evolving tensions provided as input (de Lima, Feijó, and Furtado 2019).

Our Starting Point

The evolutionary solution in (Gervás 2022) combines AoIs (see Table 1 above) using as fitness function the correct sequencing of events and acceptable occurrence of characters sharing roles across AoIs. The genetic representation employed for evolutionary construction of stories represents a narrative in terms of how the plot atoms in the AoIs are presented in the ordered sequence that constitutes the discourse of the narrative, and how the various roles for characters in the plot atoms are instantiated with identifiers for the characters in the narrative.

The fitness function that drives the evolutionary process relies on metrics for sequencing of events, and occurrence of characters sharing roles across AoIs proposed in (Gervás 2022). For each pairwise combination of AoIs the constraints on character occurrence and event sequencing are expressed in the form of constraints as shown in Table 2. The metrics assign a partial score over 100 to each sequencing constraint over events, corresponding to the number of positions that one of the elements would need to shift for the constraint to hold (normalised over the length of the sequence). Each role-sharing constraint present is scored 100 if met and 0 otherwise. The final score for a draft is computed as the weighted sum of the average value of the role-sharing constraints and the average value of the sequencing constraints. The relative weights for sequencing and role sharing constraints have been empirically set to 20 and 80.

This metric provides a progressive scoring, so that drafts where the sequencing constraints are not met are scored relative to how far they need to be modified for the constraints to be met. This allows mutations that modify the sequence in the right direction to be scored progressively higher, allowing evolution to converge towards optimal solutions.

Optimising the Evolutionary Process

A detailed study of the performance of the original algorithm lead to the identification of some shortcomings, which, when solved, lead to significant improvements in performance.

Issues with the Original Genetic Representation

The original evolutionary solution relied on a genetic representation that presented three important shortcomings. First, it represented the order in which the plot atoms appeared in the story in terms of the set of jumps to be made over the the constituent AoIs. Small changes in the set of jumps lead to very different final stories. This lead to poor exploration of the search space, because it made it difficult to explore alternatives close in the neighbourhood of given individual. Second, the genetic representation allowed jumps to be postulated even when there were no further AoIs available to jump to, having all been exhausted in prior jumps. This created situations in which different genetic representations–one that indicated a shift to another AoI at that point and one that did not–resulting in the same actual narrative. This had a negative side effect in that populations might have individuals with different genotype but equivalent phenotype. Third, the representation for the instantiations of roles from different AoIs with shared characters lead to assymmetries between different parts of the encoding vector: positions at the start of the vector had a wealth of potential candidates to be instantiated, whereas positions later in the vector could only be instantiated with characters already assigned to incompatible positions earlier. This also lead to underperformance of the evolutionary algorithm when exploring the search space.

These shortcomings went unnoticed in the early tests because it was assumed that the observed low scores were the result of incompatible restrictions for a given set of AoIs. However, more detailed consideration lead to the discovery of the negative impact of these problems in the genetic representation, which were stopping the evolutionary algorithm from reaching more desirable areas of the search space.

An Improved Genetic Representation

The original genetic representation has been replaced with a new version that resolves the observed shortcomings. It still encodes separately the order in which the plot atoms from the various AoIs appeared in the discourse and the instantiations of roles from different AoIs with shared characters of the story.

The order of appearance is now encoded as sequence of indices of the plot atoms to be included in the discourse. Each index simply indicates which plot atom from which AoI should feature next in the discourse. Mutation is now encoded as a shift of a particular index either forwards or backwards in the sequence for a number of positions chosen at random. Shifts involve skipping over plot atoms from other AoIs but they must respect the relative order of plot atoms within the same AoI. This encoding does not allow crossover operations, as cutting different representations at the same point is likely to result in drafts with either missing or redundant instances of plot atoms in some AoIs.

The instantiations of roles from different AoIs with shared characters of the story is now encoded as a set of specific data structures for encoding any variables that have a shared instantiation across pairs of AoIs. Mutation is now encoded as either adding or eliminating a connection to the data structure for a particular pair of AoIs. The choice of which pair of AoIs to consider and whether to add or eliminate are chosen at random within the bounds of available possibilities. Pairs with no connections only allow addition, pairs with all available variables already connected only allow elimination.

Although the new representation no longer allows crossover operations on the subsets of the genetic representation that encode the different aspects, a certain crossover is possible by swapping the representations of the relative order in the sequence between two different individuals to give rise to a new pair.

An example of system output encoded with the new genetic representation is shown in Table 4. This example shows together the improved genetic representation–the genotype– and the instantiation of it as story–the phenotype. Additionally it shows the intermediate data structures that translate the genetic encoding into the features that are used to construct the final draft for the narrative.

The textual rendering presented for the narrative is not intended to be the final medium for presenting it to a potential audience. Since the generation procedure described here is only concerned with the narrative structure of the plot, it is beyond the scope of the paper to evaluate or even consider aspects specific to the linguistic rendering of this content. Nevertheless a template-based transcription of the content is included to facilitate the appreciation of the narrative structure. Alternative solutions based on neural technologies, such as generative pre-trained transformers (Dale 2021), may be considered in future work.

Metric for Romantic Coherence

An undesirable feature of the early results was the fact that the resulting narratives exhibited in some cases incoherent behaviour of the characters in terms of their romantic inclinations. As many of the AoIs involve romantic relations between the characters, this often resulted in stories where characters exhibited surprisingly promiscuous behaviour, such as marrying several different characters in succession with on intervening explanation of their change of heart. These situations came about when two AoIs were combined that both included romantic relations between the characters, for instance SHIFTINGLOVE–which involves a character oscilating between two different love interests as the story evolves and deciding on one towards the end– and RELENTINGGUARDIAN–which involves a couple who wants to marry overcoming the obstacle of a guardian opposed to the match. When these two AoIs are combined, there were no safeguards in the original solution to avoid that a single character end up being matched with two different partners, one under each AoI.

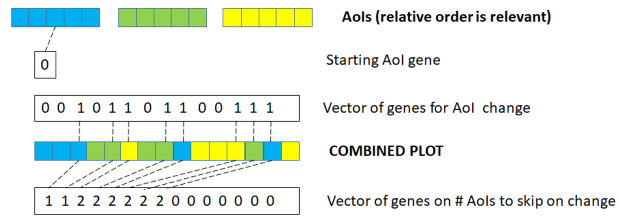

Figure 1: Original genetic representation.

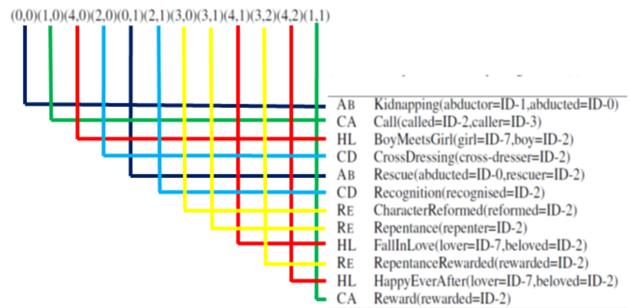

Figure 2: Improved genetic representation.

To filter out these cases, an additional component was added to the metrics that scored each character in a draft in terms of the their romantic consistency. Each character in a draft is now assigned a romantic consistency score of 100 if has at most one single romantic match, and each draft is assigned the average of the scores on romantic consistency of the characters in it. This additional metric is added to the existing fitness function, which is already computed as an average of a number of metrics on consistency over different pairs of AoIs.

Discussion

The two genetic representations considered for the discourse sequence of the story are shown in Figures 1 and 2. The original representation encoded the operations required to combine the AoIs, including a gene to indicate which AoI to start the story on, binary genes for each position in the draft to indicate whether a transition to a different AoI followed, and numerical genes to indicate how many AoIs to skip in each transition. In contrast, the new representation encodes simply the final order of the discourse.

The introduction of the new representation has lead to a significant increase in the average scores of the populations for runs on equivalent sets of inputs. The results may be compared in terms of the relative scores on quality, because both the knowledge resources and the evaluation metrics used as fitness functions remain the same. A quantitative comparison of the score for the two versions is shown in Table 3.

Whereas the results of the earlier version converged to the scores considerably below the maximum threshold, the current version reaches achieves consistently higher scores under similar circumstances. It also does frequently reach the

Version	Pop.size	Average score	Highest score
Original	100	59.2	80.1
Improved	100	67.0	94.6
Original	200	59.1	77.9
Improved	200	73.3	98.0

Table 3: Comparative scores for the evolutionary solution based on the original genetic representation and the improved version. Scores shown are averages over 10 runs of each system with the same setting: population size of 100 and 200, 20 generations, seed RAGS2RICHES, expanding with 3 additional AoIs.

maximum score. This is due to a significantly better exploration of the search space. The earlier version of the procedure must have been stuck in local optima held back by redundant encoding. This hypothesis is supported by the fact that an increase in the size of the population does not yield any significant changes in the scores. It is important to note that, while the increase in the size of the population does lead to a slight increase in the scores for the version using the improved representation, the scores for the version using the original representation are even lower than with a smaller population.

The shortcomings observed in the original genetic representation correspond to the problems of non-synonymous redundancy and low locality (Rothlauf 2006). The improved performance of the proposed solution highlights the importance for evolutionary approaches to story generation to satisfy such general requirements on genetic representations.

This improvement in the overall scores that arises from the modified genetic representation is compensated by the introduction of the additional metric on romantic consistency. With this addition, the final scores of the population recover their discriminating capability, and the results now include narratives that are coherent with respect to the romantic lives of the characters. The metric for romantic coherence is consistent with prior approaches to evaluating semantic coherence of significant events over story drafts (Gervás, Concepción, and Méndez 2022).

In general terms, the set of metrics integrated into the fitness function are defined over characteristics that are specific to the phenotype rather than the genotype of each draft. For this reason, they are applicable to any stories regardless of whether they have been produced by an evolutionary procedure or any other construction method. To apply these metrics to stories beyond the output of this system the only requirement is to provide means for the specific features being considered–narrative roles, plot-relevant events, story milestones that imply romantic commitments. . . –to be extracted from the stories to be considered. As the significance of such features for the evaluation of stories is difficult to question, the set of metrics in themselves can be considered a valuable contribution to the field.

This is especially useful in a context where the application of neural technologies has lead to a proliferation of solutions for story generation based exclusively on probabilities of word co-occurrence. Such solutions are known to be susceptible of significant improvement by means of fine tuning procedures driven by reinforcement learning based on computational reward models (Ziegler et al. 2019). If the problem of automatically extracting semantic information from text can be solved successfully, metrics such as these can prove to be valuable contributions for solutions based on large language models as support either for reward models during fine tuning or for filtering and refining outputs–in the processes known as prompt engineering.

Conclusions

The improved genetic representation proposed for evolutionary combination of plot-relevant spans of discourse solves the shortcomings observed in prior versions. The stories obtained with the enhanced version achieve significantly higher scores under the existing metrics for story quality, leading to a point where system outcomes consistently reach top scores.

As this endangers the discriminating power of the metrics for identifying higher quality stories, an extension of the metric has been proposed. The extension measures the coherence of the romantic behaviour of the characters. Under the extended set of metrics the system generates stories that have recognisable narrative structure and in which the characters are consistent in terms of their romantic relations.

The metrics on story quality proposed to inform the evolutionary fitness functions are designed to captures features relevant to the evaluation of narrative and they are independent of the genetic representation and the overall evolutionary procedure. They are therefore valuable contributions to the field of story generation in general on their own right.

As future work we intend to address issues at two different levels. In terms of richer representations of narrative structure, we will explore extensions to the construction procedure to make it capable of generating narratives as a series of connected episodes. In terms of improvements on the rendering of the narratives as text we will consider solutions for rendering the resulting plots as text that rely on generative pretrained transformers.

Acknowledgments

This paper has been partially funded by the projects CANTOR: Automated Composition of Personal Narratives as an aid for Occupational Therapy based on Reminescence, Grant. No. PID2019-108927RB-I00 (Spanish Ministry of Science and Innovation) and the ADARVE (Análisis de Datos de Realidad Virtual para Emergencias Radiológicas) Project funded by the Spanish Consejo de Seguridad Nuclear (CSN).

References

Adamowicz, E. 1998. *Surrealist collage in text and image: Dissecting the exquisite corpse*, volume 56. Cambridge University Press.

Booker, C. 2004. *The Seven Basic Plots: Why We Tell Stories*. The Seven Basic Plots: Why We Tell Stories. Continuum.

(a) genetic representation for discourse order: each pair shows (<AoI index>,<plot atom index>).

(0,0)(1,0)(4,0)(2,0)(0,1)(2,1)(3,0)(3,1)(4,1)(3,2)(4,2)(1,1)

(b) genetic representation for character instantiations: each column represents a pair of AoIs, row 1 holds indices of the AoIs, row 2 shows how many variables appear in each AoI, row 3 shows which variable from the first is connected to which of the second.

0-1	0-2	0-3	0-4	1-2	1-3	1-4	2-3	2-4	3-4
3-2	3-1	3-1	3-2	2-1	2-1	2-2	1-1	1-2	1-2
2-1			0-1	1-0	0-0		0-0	0-1	

(c) discourse plan encoded by the genes in (a).

AB	Kidnapping(abductor=ID-1,abducted=ID-0)
CA	Call(called=ID-2,caller=ID-3)
HL	BoyMeetsGirl(girl=ID-7,boy=ID-2)
CD	CrossDressing(cross-dresser=ID-2)
AB	Rescue(abducted=ID-0,rescuer=ID-2)
CD	Recognition(recognised=ID-2)
RE	CharacterReformed(reformed=ID-2)
RE	Repentance(repenter=ID-2)
HL	FallInLove(lover=ID-7,beloved=ID-2)
RE	RepentanceRewarded(rewarded=ID-2)
HL	HappyEverAfter(lover=ID-7,beloved=ID-2)
CA	Reward(rewarded=ID-2)

(d) character assignment encoded by the genes in (b).

Abduction-hero	ID-2
Abduction-victim	ID-0
Abduction-villain	ID-1
CallToAction-sender	ID-3
HappyLove-girl	ID-7
HappyLove-hero	ID-2
CrossDressing-someone	ID-2
CallToAction-hero	ID-2
Repentance-villain	ID-2

(e) template-based text rendering of the plot of the draft.

Scott kidnaps Hawa. Korr calls to action West. West meets Lilly. West dresses up as a member of the opposite sex. West rescues Hawa from Scott. West is recognised. West reforms their character. West repents. West falls in love with Lilly. West sees repentance rewarded. West lives happily ever after with Lilly. West is rewarded.

Table 4: An example of plot generated by the system by combining the following AoIs: ABDUCTION (AB), CALLTOACTION (CA), CROSSDRESSING (CD), REPENTANCE (RE), and HAPPYLOVE (HL). The procedure was initiated using ABDUCTION as a seed, to be expanded with 4 additional AoIs. The evolutionary process was run for 30 generations with a population of 500 individuals.

Cran, R. 2013. 'Everything is permitted': William Burroughs' Cut-up Novels and European Art. *Comparative American Studies An International Journal* 11(3):300–313.

Dale, R. 2021. Gpt-3: What's it good for? *Natural Language Engineering* 27(1):113–118.

de Lima, E. S.; Feijó, B.; and Furtado, A. L. 2019. Procedural generation of quests for games using genetic algorithms and automated planning. In *18th Brazilian Symposium on Computer Games and Digital Entertainment, SBGames 2019, Rio de Janeiro, Brazil, October 28-31, 2019*, 144–153. IEEE.

Forster, E. M. 1927. *Aspects of the novel*. New York: Harcourt.

Fredericks, E. M., and DeVries, B. 2021. (Genetically) Improving Novelty in Procedural Story Generation. In *2021 IEEE/ACM International Workshop on Genetic Improvement (GI)*, 39–40. IEEE.

Gervás, P.; Concepción, E.; and Méndez, G. 2022. Evolutionary construction of stories that combine several plot lines. In *Computational Intelligence in Music, Sound, Art and Design – 11th International Conference, EvoMUSART 2022*. Madrid, Spain: Springer.

Gervás, P. 2019. Generating a search space of acceptable narrative plots. In *10th International Conference on Computational Creativity (ICCC 2019)*.

Gervás, P. 2022. Evolutionary stitching of plot units with character threads. In *WIVACE 2022 XVI International Workshop on Artificial Life and Evolutionary Computation*.

Kartal, B.; Koenig, J.; and Guy, S. J. 2014. User-driven narrative variation in large story domains using monte carlo tree search. In *Procs. of AAMAS '14*, 69–76.

McIntyre, N., and Lapata, M. 2010. Plot induction and evolutionary search for story generation. In *Procs. of ACL 2010*, 1562–1572. Uppsala, Sweden: Association for Computational Linguistics.

Rothlauf, F. 2006. *Representations for genetic and evolutionary algorithms*. Springer.

Young, R. M.; Ware, S. G.; Cassell, B. A.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung* 37(1-2):41–64.

Ziegler, D. M.; Stiennon, N.; Wu, J.; Brown, T. B.; Radford, A.; Amodei, D.; Christiano, P. F.; and Irving, G. 2019. Fine-tuning language models from human preferences. *CoRR* abs/1909.08593.